

Asociación de Psicología de Puerto Rico

PO Box 363435 San Juan, Puerto Rico 00936-3435
Tel. 787.751.7100 Fax 787.758.6467
www.asppr.net E-mail: info@asppr.net

Revista Puertorriqueña de Psicología
Volumen 19, 2008

Revista Puertorriqueña de Psicología
Vol. 19, 2008

Confiabilidad de la Escala de Inteligencia Wechsler para Adultos Versión III, Puerto Rico (EIWA-III)

José I. Pons, Lizette Flores-Pabón, Leida Matías-Carrelo, Mary Rodríguez, Ernesto Rosario-Hernández
Escuela de Medicina de Ponce[1]

Juana M. Rodríguez, Laura Leticia Herrans
Universidad de Puerto Rico – Río Piedras

James Yang
Pearson / The Psychological Corporation

Resumen

En este artículo presentamos los resultados de los estudios de confiabilidad de la EIWA-III, Escala de Inteligencia Wechsler para Adultos Versión III, Puerto Rico (traducción adaptación y normalización de la WAIS III). Estos estudios psicométricos confirman la solidez de la consistencia interna y la estabilidad temporal de la EIWA-III. Concluimos que la EIWA-III es un instrumento válido y confiable para la medición de inteligencia de la población adulta de Puerto Rico.

Palabras claves: *confiabilidad, EIWA-III*

[1] Este proyecto ha sido financiado por la Escuela de Medicina de Ponce. El estudio de confiabilidad temporal se realizó como parte del proceso de normalización de la EIWA-III y sirvió además para satisfacer el requisito de disertación doctoral de la co-autora Lissette Flores-Pabón. Para información adicional sobre el artículo o sobre la prueba EIWA-III favor de comunicarse con el Dr. José I. Pons, Escuela de Medicina de Ponce, Ponce, PR. jpons@psm.edu

Nota: Este artículo fue sometido para evaluación en julio de 2008 y aceptado para publicación en noviembre de 2008.

Abstract

In this article we present the results of the reliability studies of the new EIWA III, Escala de Inteligencia Wechsler para Adultos Versión III, Puerto Rico (translation, adaptation and standardization of the WAIS III). These studies support the strength of the EIWA-III's internal consistency and temporal stability. We conclude that the EIWA-III is a valid and reliable instrument for the intellectual assessment of the adult population of Puerto Rico.

Keywords: *reliability, EIWA-III*

El proceso de estandarización de la EIWA-III comenzó en el 2003, después de concluido el estudio piloto que permitió diseñar la versión experimental de la prueba (Rodríguez, Herrans, Pons, Matías-Carrelo, Medina & Rodríguez, 2008). Para el 2007 se concluyó la fase de administración de la prueba a la muestra de normalización. Dicha muestra se diseñó siguiendo las características poblacionales de la población de adolescentes y adultos de Puerto Rico según el censo del 2000. Las características específicas de las muestras utilizadas para el estudio piloto, para la normalización, los estudios de validez y confiabilidad y los grupos clínicos se detallan en el manual de la prueba (Wechsler, 2008).

Con las puntuaciones obtenidas por la muestra de normalización se llevaron a cabo los estudios de validez y confiabilidad. Los resultados de los estudios de validez de contenido, de criterio, predictiva y de constructo confirmaron la adecuacidad psicométrica del proceso de construcción, traducción y adaptación de la prueba. Los resultados obtenidos para todos los estudios de validez apoyaron la validez del instrumento (Pons, Matías-Carrelo, M. Rodríguez, J. M. Rodríguez, Herrans, Jiménez, *et al*., 2008).

Reconociendo que la validez de una prueba no garantiza por sí sola su poder discriminativo y su calidad (Aiken, 1994; Anastasi, 1997; Herrans, 2000; Matarazzo, Wiens, Matarazzo & Manaugh, 1997), procedimos a constatar la confiabilidad del nuevo instrumento de medición. Según Anastasi y Urbina (1997), la confiabilidad de un instrumento se relaciona con la precisión, consistencia y la estabilidad de las puntuaciones que se obtienen en diversas situaciones. Esta definición implica que la confiabilidad depende de cuánto se aproximan las puntuaciones que se obtienen en la prueba a las puntuaciones *reales* que

hipotéticamente una persona puede obtener en dicha prueba bajo condiciones idóneas. Por lo tanto, partimos de la teoría clásica de las pruebas que postula que la puntuación real que una persona o grupo de personas puede producir en una prueba depende de la relación entre la puntuación observada y el error de la medida de dicha puntuación (Nunnally & Bernstein, 1995). Además, la prueba debe tener un grado adecuado de consistencia o coherencia entre sus componentes, es decir, consistencia interna. Por último, los instrumentos de medición reflejan su confiabilidad cuando permiten obtener puntuaciones razonablemente estables (o similares) en administraciones repetidas.

Para evaluar la confiabilidad de la EIWA-III llevamos a cabo dos procedimientos principales. Evaluamos la consistencia interna y la estabilidad temporal de la prueba. La primera de éstas se estudió mediante el método de dos mitades mientras que la estabilidad temporal se estudió con el método de prueba-reprueba. Se computó el Error Estándar de la Medida (EEM) de las puntuaciones de las subpruebas y de las puntuaciones compuestas de la EIWA-III para obtener información sobre el margen de error de las puntuaciones a obtenerse, basado en la teoría clásica de la prueba. El EEM permite estimar la probabilidad de que la puntuación real de la persona evaluada se encuentre entre uno, dos o tres desviaciones típicas de la puntuación obtenida. Nos dice también cuánto dista una puntuación observada de la puntuación hipotetizada con la real y que se supone obtenga la persona evaluada, según ya indicado. La totalidad de estas estadísticas se han incluido en el manual de la prueba (Wechsler, 2008) y se utilizaron como base para computar los intervalos de confianza de los cocientes intelectuales y los índices de la EIWA-III. En este artículo sólo presentamos la información necesaria para establecer y permitir a la audiencia evaluar la confiabilidad de la escala.

Niveles de Confiabilidad de las Escalas Wechsler

Las escalas de inteligencia Wechsler han gozado de excelentes niveles de consistencia interna y de estabilidad temporal desde la publicación de sus primeras versiones. Es decir, los índices promedios de confiabilidad de consistencia interna usualmente se encuentran alrededor de .90. Estos niveles de confiabilidad permiten hacer inferencias y predicciones a nivel individual a base de los resultados obtenidos (Herrans, 2000; Nunnally, 1978). Niveles de confiabilidad como éstos se encontraron en la Escala de Inteligencia Wechsler para Adultos (EIWA), la que fue normalizada para Puerto Rico durante la década de los sesenta (Wechsler, 1968) siendo la versión al español de la segunda escala diseñada por David Wechsler, la *Wechsler Intelligence Scale*

for Adults (Wechsler, 1955). La EIWA obtuvo un coeficiente de consistencia interna (mediante división de mitades) de .97 para la Escala Verbal, de .95 para la Escala de Manipulación y de .98 para la Escala Total. La segunda escala de inteligencia diseñada para niños por David Wechsler (Wechsler, 1974), la *Wechsler Intelligence Scale for Children-Revised*, fue normalizada para Puerto Rico por Herrans y Rodríguez (Wechsler, 1992) y denominada Escala de Inteligencia Wechsler para Niños Revisada (EIWN-R PR). Los índices de consistencia interna de la EIWN-R cumplen con las expectativas psicométricas. El índice del CIV es de .92, y el del CIT de .94. El índice de consistencia interna del CIE de la EIWN-R PR es de .88, lo que se considera también, muy bueno.

La más reciente versión de las escalas de inteligencia Wechsler para adultos, la *Wechsler Adult Intelligence Scale-III* (WAIS-III), obtuvo coeficientes de consistencia interna de .97 para la Escala Verbal, de .94 para la Escala de Ejecución y de .98 para la Escala Total, todos sobrepasando los niveles estimados como excelentes para una escala de este tipo. Los coeficientes de los índices son igualmente excelentes fluctuando entre .93 (IOP) y .96 (ICV). El Índice de Velocidad de Procesamiento (IVP) es de .88 catalogado como muy bueno. Los coeficientes de las subpruebas de la WAIS-III fluctúan entre .70 para Composición de Objetos y .93 para Vocabulario. La subprueba Composición de Objetos, no fue incluida en la EIWA-III ni será incluida en la nueva versión de las escalas Wechsler para adultos, la WAIS-IV (L. Weiss -*Pearson/The Psychological Corporation*, comunicación personal, 20 de marzo de 2004).

Los coeficientes de confiabilidad prueba-reprueba de las escalas Wechsler han sido y son igualmente sólidos. La WISC-III obtuvo coeficientes de estabilidad temporal para todas sus subpruebas, índices y CI entre .86 y .94. La versión de la WISC-R normalizada para Puerto Rico por Herrans y Rodríguez (Wechsler, 1992), la EIWN-R PR, obtuvo coeficientes de confiabilidad muy similares que fluctuaron entre .88 a .94. Las escalas Wechsler para pre-escolares también han reflejado altos niveles de estabilidad temporal comenzando con la WPSSI original que obtuvo coeficientes entre .87 y .91 (Wechsler, 1967). La versión revisada de esta escala, la WPSSI-R, obtuvo coeficientes promedio para las puntuaciones compuestas de CI entre .92 y .96, los que se consideran excelentes indicadores de estabilidad temporal (Wechsler, 1989).

La estabilidad temporal de la WAIS-III es igualmente adecuada (Wechsler, 1997). El coeficiente de estabilidad temporal del CIV de la WAIS-III es .96, el del CIE es de .91 y el del CIT es de .96. Para los

índices se obtuvieron coeficientes que fluctúan entre .88 para el IOP y .95 para el ICV.

Estudio de Consistencia Interna de la EIWA-III

Mediante el estudio de consistencia interna de la EIWA-III se intentó determinar la consistencia con la que los reactivos del instrumento miden las características o rasgos latentes que intentan medir. Evaluamos los reactivos de las subpruebas para obtener esta información y subsiguientemente calculamos los índices de las puntuaciones compuestas. El método que con más frecuencia se utiliza para evaluar la consistencia interna de las escalas de inteligencia, especialmente las escalas Wechsler, es el de división de mitades. Para implementar esta técnica utilizamos las puntuaciones obtenidas en la prueba por la muestra de normalización, la que constó en este caso de 330 participantes. La descripción de las muestras utilizadas para el proceso de estandarización de la EIWA-III se describe en el manual de la prueba (Wechsler, 2008) y en el artículo de Rodríguez, Herrans, Pons, Matías-Carrelo, Medina y Rodríguez (2008), en este volumen.

El procedimiento de división de mitades permite evaluar la confiabilidad de consistencia interna de una prueba a base de una sola administración lo que evita la posible alteración de los resultados como consecuencia de la práctica o del factor tiempo. Para implementar este método organizamos los reactivos de las subpruebas en rangos tomando como base los niveles de dificultad obtenidos mediante la Teoría de la Respuesta a Ítemes (IRT por sus siglas en inglés). Dividimos los reactivos de las subpruebas en pares y nones para crear dos mitades de cada subprueba. Obtuvimos el coeficiente de confiabilidad de cada subprueba al correlacionar ambas mitades mediante la prueba de correlación Pearson. Sin embargo, cuando se correlacionan las mitades de una subprueba se disminuye el coeficiente de correlación en tanto el número de reactivos que se correlacionan ha sido disminuido. Para corregir el efecto de esta disminución se utilizó la fórmula Spearman-Brown (Crocker & Algina, 1986; Li, Rosenthal, & Rubin, 1996). La fórmula de correlación Spearman-Brown eleva la correlación de las dos mitades de las subpruebas al nivel que se obtendría si la correlación se hubiese realizado con la totalidad de los reactivos. Este procedimiento no aplica a la evaluación de confiabilidad de las subpruebas que miden velocidad de procesamiento (Identificación de Símbolo, Dígito Símbolo-Clave) para lo que se utiliza el método de prueba-reprueba.

En la Tabla 1 observamos un número mayor de subpruebas verbales con coeficientes superiores a .90 que entre las de ejecución.

TABLA 1
Coeficientes de Consistencia Interna de las Subpruebas de la EIWA-III por Grupos de Edades

Sub-prueba	*Grupos de Edades*										
	16-17	*18-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*	*50-54*	*55-59*	*60-64*
Voc	.83	.88	.89	.90	.90	.96	.87	.91	.93	.93	.93
Sem	.81	.84	.91	.87	.93	.93	.89	.85	.90	.90	.86
Arit	.91	.74	.91	.80	.88	.90	.81	.91	.87	.87	.80
RetD	.77	.69	.88	.80	.80	.81	.86	.85	.74	.74	.68
Inf	.90	.86	.92	.92	.96	.94	.89	.89	.95	.95	.92
Comp	.78	.80	.91	.88	.84	.88	.91	.69	.86	.86	.85
SecLN	.91	.68	.92	.75	.74	.81	.77	.86	.77	.77	.84
FigI	.84	.67	.88	.88	.94	.93	.85	.91	.92	.92	.87
DígSC	.81	.81	.81	.81	.81	.81	.81	.81	.81	.81	.81
DisB	.84	.83	.92	.89	.89	.89	.80	.84	.87	.87	.86
MatR	.92	.79	.93	.87	.93	.81	.90	.95	.89	.89	.94
ArrD	.61	.67	.63	.78	.57	.89	.78	.69	.90	.90	.76
IdS[6]	.73	.73	.73	.73	.73	.73	.73	.73	.73	.73	.73

Nota: Los coeficientes de confiabilidad para Identificación de Símbolos y Dígito Símbolo-Clave se basan en el método de prueba-reprueba. Voc = Vocabulario, Sem = Semejanza, Arit = Aritmética, RetD = Retención de Dígitos, Info = Información, Comp = Comprensión, SecLN = Secuencia de Letras y Números, FigI = Figuras Incompletas, DígSc = Dígito Símbolo-Clave, DisB = Diseño con Bloques, MatR = Matriz de Razonamiento, ArrD = Arreglo de Dibujos, IdS = Identificación con Símbolos.

Específicamente, las subpruebas de Información, Vocabulario y Aritmética contienen la mayor incidencia entre todos los grupos de edades de índices superiores a .90. De las subpruebas de ejecución, Figuras Incompletas presenta la mayor incidencia de índices mayores de .90 entre los diversos grupos de edades. En cinco de las subpruebas (FI, Comp, SecLN, RD y AD), principalmente en el grupo de 18-19 años, se obtienen coeficientes menores de .70, entre .61 y .69, los que se consideran adecuados, aunque de forma leve. Los índices más bajos de confiabilidad se obtienen en la subprueba de Arreglo de Dibujo, el más bajo, .57, en el nivel de edad de 30 a 34 años. Dos de las subpruebas obtuvieron el mismo coeficiente para todos los rangos de edades, .81 y .73. Son éstas las subpruebas de Dígito Símbolo-Clave e

TABLA 2
Coeficientes de Consistencia Interna de los CI de la EIWA-III por Grupos de Edades

CI	*Grupos de Edades*										
	16-17	*18-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*	*50-54*	*55-59*	*60-64*
CIV	.96	.94	.98	.97	.97	.98	.97	.96	.97	.97	.96
CIE	.91	.90	.94	.94	.95	.96	.95	.95	.97	.97	.95
CIT	.97	.96	.98	.98	.98	.98	.98	.98	.98	.98	.98

TABLA 3
Coeficientes de Consistencia Interna de los Índices de la EIWA-III por Grupos de Edades

Índices	*Grupos de Edades*										
	16-17	*18-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*	*50-54*	*55-59*	*60-64*
ICV	.93	.94	.96	.96	.97	.98	.96	.95	.97	.97	.96
IOP	.93	.87	.96	.95	.97	.94	.93	.95	.95	.95	.94
IMT	.94	.83	.96	.87	.92	.93	.92	.94	.91	.91	.89
IVP	.85	.84	.86	.85	.86	.86	.88	.83	.87	.87	.87

Identificación de Símbolos, que según hemos indicado se evalúan de modo más adecuado mediante el método de prueba-reprueba. A continuación se presentan las Tablas 2 y 3 con los coeficientes de consistencia interna para los CI y los Índices.

Existen diferencias entre los niveles de consistencia interna obtenidos por las subpruebas al nivel individual versus las puntuaciones compuestas. Los niveles de consistencia interna en las puntuaciones compuestas se basan en múltiples subpruebas y se espera que obtengan coeficientes más consistentes entre grupos de edades y entre los diferentes Índices, los que alcanzan en su mayoría coeficientes mayores de .90. De hecho, como podemos ver en la Tabla 2, los coeficientes de los CI presentan poca variabilidad y se ubican para todos los grupos de edades entre .90 y .98 lo que se considera robusto y altamente deseable. De forma parecida, la Tabla 3 nos muestra la relativamente poca variabilidad entre los coeficientes de confiabilidad interna, que oscilan entre .83 y .98, para todos los Índices en todos los grupos de edad.

La Tabla 4 presenta el EEM obtenido para los CI y la Tabla 5 los

TABLA 4
Error Estándar de la Medida de los CI de la EIWA-III por Grupos de Edades

	Grupos de Edades										
CI	*16-17*	*18-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*	*50-54*	*55-59*	*60-64*
CIV	3.00	3.67	2.12	2.60	2.60	2.12	2.60	3.00	2.12	2.60	3.00
CIE	4.50	4.74	3.67	3.67	3.35	3.00	3.35	3.35	3.00	2.60	3.35
CIT	2.60	3.00	2.12	2.12	2.12	2.12	2.12	2.12	2.12	2.12	2.12

TABLA 5
Error Estándar de la Medida de los Índices de la EIWA-III por Grupos de Edades

	Grupos de Edades										
Índices	*16-17*	*18-19*	*20-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*	*50-54*	*55-59*	*60-64*
ICV	3.97	3.67	3.00	3.00	2.60	2.12	3.00	3.35	2.60	2.60	3.00
IOP	3.97	5.41	3.00	3.35	2.60	3.67	3.97	3.35	3.00	3.35	3.67
IMT	3.67	6.18	3.00	5.41	4.24	3.97	4.24	3.67	3.00	4.50	4.97
IVP	5.81	6.00	5.61	5.81	5.61	5.61	5.20	6.18	5.61	5.41	5.41

Nota: 1. El error estándar de la medida se informa en unidades de puntuaciones a escala y en unidades de CI/Índices para los CI y para las puntuaciones de los índices. Los coeficientes de confiabilidad y las desviaciones típicas de la población (3 para las subpruebas y 15 para las puntuaciones de CI y de los Índices) se usaron para calcular el error estándar de la medida.

EEM computados de los Índices. La Tabla 6 contiene el nivel de EEM asociado al coeficiente de confiabilidad de cada subprueba. En el manual técnico y de administración de la EIWA-III (Wechsler, 2008) se ilustra el uso de esta información para obtener los intervalos de confianza de los CI y de los Índices de la prueba. El EEM, al ser convertido en intervalos de confianza, provee otro medio de expresar la precisión de las puntuaciones de las pruebas.

El EEM provee una impresión o un estimado del error presente en la puntuación observada que se obtiene de la prueba. Según hemos indicado anteriormente, el nivel de *error* disminuye la probabilidad de que la puntuación observada se aproxime a la puntuación *real* que se espera que obtenga la persona o el grupo evaluado con la prueba. Por lo tanto, existe una relación inversa entre EEM y confiabilidad. A menor la puntuación que se obtiene de EEM, mayor es la confiabili-

dad de la prueba (o subprueba en el caso de las escalas Wechsler) lo que permite a las personas que utilizan la prueba tener mayor nivel de confianza en los resultados que se obtienen.

Los índices de consistencia interna que presentamos en la Tabla 6 se obtienen del promedio de los Índices de todos los rangos de edades de cada subprueba. De igual modo, los EEM que se presentan en esta tabla son el promedio de los EEM de todos los grupos de edades. Observamos que los EEM que se presentan en esta tabla fluctúan entre 0.87 y 1.56. Estos niveles concuerdan con lo que se obtiene usualmente por este tipo de escala. En su mayoría los EEM obtenidos para la EIWA-III son relativamente bajos ya que la desviación estándar de la escala en la cual se convierten estas puntuaciones a escala es de 3 puntos. La desviación estándar de las escalas de los Índices y los CI es de 15 puntos, motivo por el cual se observarán puntuaciones de EEM

TABLA 6
Índice de Confiabilidad y EEM Obtenido para las Subpruebas de la EIWA-III

Subpruebas	*Confiabilidad*	*EEM*
Vocabulario	.91	0.94
Semejanzas	.89	1.02
Aritmética	.87	1.13
Retención Dígitos	.80	1.36
Información	.92	0.87
Comprensión	.85	1.19
Secuencia de Letras y Números	.83	1.30
Figuras Incompletas	.88	1.08
Dígito Símbolo Clave	.81	1.31
Diseño con Bloques	.87	1.11
Matriz de Razonamiento	.91	0.96
Arreglo de Dibujos	.76	1.54
Identificación de Símbolo	.73	1.56

necesariamente mayores. Estas diferencias psicométricas de las escalas no permiten comparar los EEM de las subpruebas con los de las puntuaciones compuestas.

La Tabla 7 presenta el EEM asociado a cada índice de confiabilidad de los CI. El Índice de Memoria de Trabajo y el de Velocidad de Procesamiento presentan los mayores valores de EEM. Esto podría ser el resultado de la diferencia en la naturaleza de las subpruebas de estos Índices en comparación con las de los Índices de comprensión verbal y de organización perceptual. Para el IVP se utilizan tareas manuales que están condicionadas por tiempo. El IMT está compuesto de subpruebas basadas en tareas de memoria verbal que no requieren tanto de ligereza, pero sí de precisión. Por lo tanto, el lapso de tiempo no es tan importante en estas subpruebas. Además, estas subpruebas contienen muchos menos reactivos que las subpruebas que componen los Índices de Comprensión Verbal y de Organización Perceptual. Se anticipa que estas diferencias en construcción de prueba se manifiesten mediante diferencias en los niveles de consistencia interna en estos índices, según ocurre en la EIWA-III.

TABLA 7

Índices de Confiabilidad y EEM Obtenido para las Puntuaciones Compuestas de la EIWA-III

Puntuaciones Compuestas	*Confiabilidad*	*EEM*
Cociente Intelectual Verbal	.97	2.71
Cociente Intelectual de Ejecución	.95	3.56
Cociente Intelectual Total	.98	2.26
Índice de Comprensión Verbal	.96	3.03
Índice de Organización Perceptual	.94	3.65
Índice de Memoria de Trabajo	.92	4.36
Índice de Velocidad de Procesamiento	.86	5.67

Confiabilidad de Estabilidad Temporal

La confiabilidad temporal se refiere a la consistencia con la que una prueba logra medir a través del tiempo las características o rasgos que evalúa. Se espera que los coeficientes de confiabilidad se encuentren entre .60 y .90 para que se pueda concluir que los niveles de con-

fiabilidad de la prueba permiten hacer predicciones certeras. Según ya hemos indicado, las escalas Wechsler obtienen tradicionalmente altos niveles de confiabilidad temporal. Lo que implica que se pueden esperar puntuaciones similares en administraciones repetidas, al tomar en cuenta las variables que afectan la estabilidad de las pruebas como por ejemplo, el factor de práctica y el intervalo de tiempo que transcurre entre las administraciones.

El efecto de la práctica se refiere específicamente a la tendencia de las puntuaciones a aumentar cuando se administra la prueba por segunda vez a una misma persona o grupo de personas. Varios estudios sugieren que a más corto el lapso de tiempo entre prueba y reprueba, más significativo es el efecto de la práctica (Thompson & Molly, 1993). Este efecto de práctica se reduce a medida que aumenta el tiempo trascurrido entre prueba y reprueba. Muchos autores advierten que este efecto de la práctica puede ser en ocasiones confundido con cambios en la condición de un individuo lo que podría conducir a interpretaciones erróneas de las puntuaciones obtenidas (Matarazzo, *et al*., 1997; Thompson & Molly, 1993).

Thompson y Molly (1993) evaluaron la estabilidad temporal de la WAIS-R a través de intervalos de tiempo que fluctuaron entre 3 y 8 meses. Obtuvieron coeficientes estables con adolescentes de 16 años donde los coeficientes de los CI fueron altos tanto para los intervalos cortos de reprueba de 3 meses (entre 3.6 y 10.7 puntos), como para los largos de 8 meses (entre 7.3 y 12.2 puntos). En ambos casos el mayor aumento de puntos se observó en la Escala de Ejecución. Los coeficientes de estabilidad obtenidos para la WAIS-R mediante este estudio fluctuaron entre .69 y .88 para el intervalo corto y entre .71 a .90 para el intervalo largo. Las diferencias entre estos valores no se encontraron estadísticamente significativas.

Raguet, Campbell, Berry, Schmitt y Smith (1996) hicieron estudios para determinar la estabilidad temporal (luego de un periodo de cerca de un año), en una población de viejos en los resultados de la WAIS-R, entre otros instrumentos. El estudio evidenció una confiabilidad temporal mayor a .90 para la WAIS-R en esa población.

Haynes y Howard (1986) indican que a mayor el lapso de tiempo entre la administración inicial y la segunda administración, menor es el coeficiente de confiabilidad. También indican que los estudios han demostrado que el efecto de la práctica es más aparente después de varios meses que después de dos o más años. Kaufman (1990) encuentra que personas normales al igual que poblaciones clínicas usualmente obtienen ganancias de aproximadamente tres puntos en la Escala Verbal, nueve puntos en la Escala de Ejecución y seis puntos de ganancia en la Escala Total. Estos hallazgos de Kaufman (1990) coinciden

con los estudios de estabilidad temporal realizados por *The Psychological Corporation* como parte del proceso de estandarización de la WAIS-III, según reportado en el manual técnico de la prueba (Wechsler, 1997). El promedio del intervalo de tiempo transcurrido entre la prueba-reprueba de la WAIS-III fue de 34.6 días y las ganancias en puntuaciones obtenidas entre los diferentes componentes de la prueba fluctuaron entre 2.0 a 8.3 puntos.

Siders, Kaufman y Reynolds (2006) estudiaron la naturaleza del efecto de la práctica en la WISC-III. El intervalo entre pruebas fluctuó entre 11 y 14 días. Las ganancias obtenidas para la segunda administración fueron entre 1.10 y 13.78 puntos.

Muestra del Estudio de Estabilidad Temporal de la EIWA-III

Para el estudio de estabilidad temporal se administró la EIWA-III a una muestra de 41 participantes de los cuales 53.7% son varones y 46.3% féminas entre las edades de 17 a 64 años. Se administró la EIWA-III a cada participante en dos ocasiones. La segunda administración se realizó entre 2 y 14 semanas. La muestra contiene residentes de diferentes sectores de la isla de Puerto Rico promoviendo representación de la mayor parte de la población. Las características de la muestra con relación a las variables de edad y niveles educativos se detallan en las Tablas 8 y 9.

TABLA 8

Distribución de Edades de la Muestra del Estudio de Estabilidad Temporal

Grupo De Edades	*Frecuencia*	*Porcentaje*	*Frecuencia Acumulativa*	*Porcentaje Acumulativo*
16 – 17	2	4.88	2	4.88
18 – 19	6	14.63	8	19.51
20 – 24	5	12.20	13	31.71
25 – 29	1	2.44	14	34.15
35 – 39	5	12.20	19	46.34
40 – 44	6	14.63	25	60.98
45 – 49	5	12.20	30	73.17
50 – 54	1	2.44	31	75.61
55 – 59	3	7.32	34	82.93
60 – 64	7	17.07	41	100.00

TABLA 9
Niveles Educativos de la Muestra del Estudio de Estabilidad Temporal de la EIWA-III

Nivel Educativo	*Frecuencia*	*Porcentaje*	*Frecuencia Acumulativa*	*Porcentaje Acumulativo*
1	1	2.44	1	2.44
2	1	2.44	2	4.88
3	9	21.95	11	26.83
4	8	19.51	19	46.34
5	22	53.66	41	100.00

Nota: Los niveles educativos utilizados corresponden a: Nivel 1 = 0 a 8 años, Nivel 2 = 9 a 11 años, Nivel 3 = 12 años, Nivel 4 = 13 a 15 años, y Nivel 5 = 16 años o más.

TABLA 10
Intervalo de Tiempo por Días entre Administraciones de la EIWA-III y Frecuencia y Porcentaje de Participantes que Ubican en Cada Intervalo

Intervalo	*Frecuencia*	*Porcentaje*
14 – 24	8	19.52
26 – 46	13	31.72
48 – 58	6	14.64
60 – 75	6	14.64
83 –108	8	19.52

El promedio de tiempo de la segunda administración de la EIWA-III fue de 50.5 días. La Tabla 10 contiene la frecuencia de participantes que se ubican en los diferentes intervalos de tiempo entre pruebas.

En la Tabla 11 se presenta la puntuación promedio y la desviación estándar obtenida por cada grupo de edad de la muestra en cada subprueba durante las dos administraciones del proceso de prueba-reprueba. Se detallan además las correlaciones obtenidas y el tamaño del efecto entre los coeficientes que se obtienen para las correlaciones corregidas para cada variable.

Según podemos observar en la Tabla 11 las correlaciones corregidas para las subpruebas verbales fluctúan entre .71 y .91. Las correlaciones para las subpruebas de ejecución fluctúan entre .73 y .81. En esta tabla podemos observar que los coeficientes de estabilidad se man-

TABLA 11

Coeficientes de Estabilidad Corregidos y el Tamaño del Efecto Computado para las Puntuaciones a Escala Obtenidas de las Subpruebas de la EIWA-III Mediante Prueba-Reprueba

Subprueba CI, Índices	*1ra. Administración*		*2da. Administración*				*Tamaño*
	$\bar{X}$	*Desv. Est.*	$\bar{X}$	*Desv. Est.*	*r*	*r Corregido*[1]	*del Efecto*[2]
Voc	11.24	2.47	11.46	2.30	.87	.91	0.09
Sem	11.34	2.51	11.88	2.40	.84	.89	0.21
Arit	11.05	2.88	11.55	2.65	.79	.81	0.18
RetD	10.63	2.63	11.00	3.06	.65	.73	0.13
Inf	10.83	2.48	11.44	2.67	.87	.91	0.24
Comp	11.55	2.50	11.45	2.47	.70	.79	0.04
SecLN	10.90	2.71	11.41	2.90	.64	.71	0.18
FigI	10.41	3.04	12.10	3.21	.81	.80	0.54
DígSC	11.13	3.38	12.80	3.53	.85	.81	0.48
DisB	10.08	3.47	10.98	3.22	.74	.65	0.27
MatR	11.93	2.79	12.15	2.92	.71	.75	0.08
ArrD	10.32	2.65	11.98	2.70	.72	.78	0.62
IdS	10.88	3.32	11.71	3.52	.78	.73	0.24

Nota: 1. Las correlaciones se corrigieron para la variabilidad de la muestra de estandarización (Allen & Yen, 1979; Magnusson, 1967).
2. Se utilizó la fórmula de Cohen (1996) para computar el tamaño del efecto.

tienen entre los niveles aceptables y excelentes para la mayoría de las subpruebas. Las subpruebas de Vocabulario e Información obtuvieron los coeficientes de confiabilidad más altos, .91 para ambas subpruebas. El coeficiente de confiabilidad más bajo lo obtuvo la subprueba de Diseño con Bloques (.65). Estos coeficientes sugieren que las subpruebas de la EIWA-III poseen estabilidad adecuada a través del tiempo. La Tabla 12 contiene el mismo tipo de información que se presenta en la tabla anterior, pero para los Índices y los CI.

Los coeficientes de estabilidad del CIV, CIT y el ICV son excelentes, entre .90 y .95. El análisis del efecto de práctica sobre las puntuaciones de los compuestos indica que para el CIV el impacto no es alto ya que las puntuaciones obtenidas difieren por sólo 2.41 al ser el

TABLA 12

Coeficiente de Estabilidad Corregidos y el Tamaño del Efecto Computado Para las Puntuaciones Obtenidas de los Índices y los CI de la EIWA-III Mediante Prueba-Reprueba

Subprueba CI Índices	1ra Administración		2da Administración				
	$\overline{X}$	Desv. Est.	$\overline{X}$	Desv. Est.	r	r Corregido[1]	Tamaño del Efecto[2]
CIV	106.05	12.59	108.46	12.77	.88	.92	0.19
CIE	104.56	14.64	112.76	15.14	.83	.84	0.55
CIT	105.51	13.23	110.68	13.49	.87	.90	0.39
ICV	105.73	11.70	108.34	12.15	.92	.95	0.22
IOP	104.50	14.97	109.33	14.86	.79	.79	0.32
IMT	104.54	13.51	107.44	13.87	.77	.81	0.21
IVP	105.46	17.38	113.41	19.63	.86	.81	0.43

Nota: 1. Las correlaciones se corrigieron para la variabilidad de la muestra de estandarización (Allen & Yen, 1979; Magnusson, 1967).
2. Se utilizó la fórmula de Cohen (1996) para computar el tamaño del efecto.

CIV de la primera administración de 106.05 y de 108.46 para la segunda administración. Esta diferencia es cónsona con lo que se obtuvo para el estudio de prueba-reprueba de la WAIS-III (Wechsler, 1997). Para todos los niveles de edades la ganancia en puntuaciones del CIV a consecuencia del efecto de práctica fue menor de tres puntos, de 2.5 a 3.2. La ganancia obtenida en el CIE es de 8.8, lo que es cónsono también con las puntuaciones obtenidas por los diversos grupos de edades de la muestra de la WAIS-III, entre 2.5 y 8.3. La media del CIT obtenido por la muestra de la EIWA-III fue de 104.56 para la primera administración y de 112.76 para la segunda, una ganancia prácticamente idéntica a la que se obtuvo para el CIE de la WAIS-III en los grupos de edades 16-29 y 30-54.

El IVP promedio obtenido para la primera administración fue de 105.46 y de 113.41 para la segunda administración. La ganancia fue de 7.95, prácticamente 8 puntos y similar al CIE. El IVP está compuesto

por solamente dos subpruebas lo que podría tener un efecto adverso en términos de práctica. Más aún, se reconoce que las subpruebas de ejecución de la Wechsler tienen ganancias mayores debido a reprueba que las subpruebas verbales.

Evaluación del Tamaño del Efecto: Estudio de Confiabilidad Temporal

El Tamaño del Efecto (TE) es un índice estadístico muy útil que tiene múltiples aplicaciones. El TE provee información sobre la magnitud o la fortaleza de la diferencia entre las medias de los grupos, lo que permite una interpretación más detallada de las puntuaciones que se obtienen de los grupos y de la significancia estadística "*p*" (Cohen, 1996). Mediante la significancia estadística se comunica si existe una diferencia entre parámetros, o si existe una relación o asociación de importancia. El TE informa el tamaño de dicha relación o asociación. Frecuentemente se utiliza el TE para evaluar el efecto de una variable independiente sobre una dependiente, o sea, el efecto del tratamiento experimental, y por lo tanto, nos permite "cuantificar el punto hasta el cuál la estadística de una muestra es divergente de la hipótesis nula" (Thompson, 2006, p. 172).

Estadísticamente TE es frecuentemente denominado "Cohen's *d*" (Thompson, 2006) y suele ser abreviado "*d*", ya que el cómputo se basa en la *diferencia* entre las medias de las puntuaciones de los grupos comparados. La fórmula de TE utilizada para el estudio de prueba-reprueba de la EIWA-III es la de Cohen (1996). Esta fórmula indica que la diferencia estándar (*d*) que constituye TE, es la diferencia de las medias de las puntuaciones de los dos grupos dividida entre la raíz cuadrada de la desviación típica conglomerada (*pooled* en inglés). Se informa que TE es pequeño cuando *d* es igual a 0.2. Cuando *d* es igual a 0.5, se considera moderado y cuando es igual o mayor a 0.8, se denomina grande. Cuando el TE es de 0.8 la media del grupo tratado se encuentra en la percentila 78 en comparación con el grupo no tratado. Cuando el TE es de 0.0 las puntuaciones del grupo tratado se encuentran traslapadas o superpuestas con las medias de puntuaciones del grupo no tratado, lo que indica que sus distribuciones son iguales. Por lo tanto, un TE alto indica que no hay traslapo entre las distribuciones de los dos grupos, por lo que se entiende que sus distribuciones son diferentes. Un traslapo bajo, menor de 0.5 o mejor aún, de 0.2 o menor, indica que las puntuaciones obtenidas por los grupos son muy similares. En términos porcentuales, un TE de 0.2 equivale a un traslapo de 14% de las distribuciones de los grupos, mientras que un TE mediano de 0.5 equivale a un traslapo de 33%. Para un TE de 0.8 el traslapo es

de 47% (Cohen, 1988).

En los estudios correlacionales se puede utilizar el TE correlacional lo que permite apreciar si los dos grupos son similares o diferentes. Los coeficientes de correlación también proveen información sobre tamaño del efecto. Sin embargo, se está utilizando el TE con estudios de confiabilidad temporal para evaluar la hipótesis de que las medias de las puntuaciones son suficientemente similares como para confirmar que la prueba mide los rasgos psicológicos consistentemente a través del tiempo.

El TE computado para el CIV de la EIWA-III es 0.19, representando aproximadamente 14% de traslapo y considerado muy bajo. El TE del CIE es 0.55 con aproximadamente 33% de traslapo, considerado moderado. El TE del CIT es 0.39 con menos de 27% de traslapo considerándose moderadamente bajo. Según expuesto, el efecto de la práctica suele ser mayor en las subpruebas de ejecución que en las verbales lo que nos explica el TE que obtuvimos para el CIE. El TE de 0.19 confirma que el coeficiente de correlación obtenido para este CI (.92) se basa en una correlación fuerte entre las dos administraciones. El TE del CIT es más alto que el del CIV, pero es aún relativamente bajo sugiriendo que la diferencia entre las puntuaciones del grupo en las dos administraciones no son tan fuertes o marcadas.

El TE del ICV es de 0.22, el del IOP es de 0.32, el del IMT es de 0.21 y el del IVP es de 0.43. Por lo tanto, la diferencia en puntuaciones en el ICV y el IMT es baja y refleja la poca variabilidad en la ejecución entre administraciones. Seis de las 13 subpruebas de la EIWA-III obtuvieron TE menores de 0.2, y 11 subpruebas menores de 0.5. Solamente dos subpruebas entran al rango de moderadas al sobrepasar 0.5, alcanzando una de éstas el nivel de 0.62, también considerado moderado. Por lo tanto, constatamos que las distribuciones de puntuaciones de la mayoría de las subpruebas son desde muy similares hasta moderadamente similares (entre – 0.04 y .48) y esta información nos ofrece evidencia adicional sobre la estabilidad temporal de la escala.

Conclusión

La EIWA-III presenta niveles de confiabilidad cónsonos con lo esperado para una escala de medición intelectual. La prueba tiene excelentes niveles de consistencia interna y las puntuaciones que se obtienen en una administración tienden a correlacionar con las que se obtienen durante administraciones subsiguientes a corto plazo (menos de tres meses). El EEM que se obtiene de las subpruebas y de las puntua-

ciones compuestas es cónsono con lo que tradicionalmente se obtiene de escalas de inteligencia de excelencia comprobada. Concluimos que la EIWA-III es un instrumento adecuado para sustituir la escala que hasta el presente se utiliza para evaluar la inteligencia de adultos puertorriqueños.

REFERENCIAS

Aiken, L. (1994). *Psychological testing and assessment*. Massachusetts: Ally & Bacon.

Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterrey, CA: Brooks/Cole.

Anastasi, A. & Urbina, S. (1998). *Test psicológicos*. México: Prentice Hall.

Anderson, P., Cronin, M., & Kazmierski, S. (1989). WISC-R stability and re-evaluation of learning-disabled students. *Journal of Clinical Psychology*. *45*, 941-944.

Axelrod, B. N., Brines, B., & Rapport, L. J. (1997). Estimating full scale IQ while minimizing the effect of practice. *Assessment, 4*, 221-227.

Barbero, M. I., Vilá, A., & Holgado, F. P. (2002). Cursos de psicometría para el año 2003-04. Accedido el 15 de septiembre de 2005, de http://www.uned.es/psico-3-psicometria/

Basso, M. R., Carona, F., Francine, D., Lowery, N. & Axelrod, B. N. (2002). Practice effects on the WAIS-III across 3 and 6 months intervals. *Clinical Neuropsychologist*, *16*, 57-63.

Bird, C. M., Papadopoulou, K., Ricciardelli, P., Rossor M. N., & Cipolotti, L. (2003). Test retest reliability, practice effects and reliable change indices for the recognition memory test. *British Journal of Clinical Psychology, 42*, 407-425.

Boone, D. E. (1992). Reliability of the WAIS – R with psychiatric inpatients. *Journal of Clinical Psychology, 48*, 72-76.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum .

Cohen, J. (1996). *Explaining psychological statistics*. Pacific Grove, CA: Brooks/Cole.

Dacey, C. M., Nelson III, W. M., & Stoeckel, J. (1999). Reliability,

criterion-related validity and qualitative comments of the fourth edition of the Stanford – Binet Intelligence Scale with a young adult population with intellectual disability. *Journal of Intellectual Disability Research, 43,* 179-184.

Deary, I., Austin, E. & Caryl, P. (2000). Testing versus understanding human intelligence. *Psychology, Public Policy, and Law. 6*, 180-190.

Escorial, S., Rebollo, I., García, L., Colom, R., Abad, F. & Espinosa, M. (2003). Las aptitudes que se asocian al declive de la inteligencia: evidencias a partir del WAIS-III. *Psicothema, 15*, 19-22.

Groth–Marnat, G. (1999). *Handbook of Psychological assessment: With WAIS–III Supplement*. New York: John Wiley & Sons.

Haynes, J. & Howard, R. (1986). Stability of WISC-R scores in a juvenile forensic sample. *Journal of Clinical Psychology*, *42*, 534-536.

Herrans, L. L. (2000). *Psicología y medición. El desarrollo de pruebas psicológicas en Puerto Rico*. México: Mc Graw Hill.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.

Lemay, S., Bédard, M., Rouleau, I., & Tremblay, P. G. (2004). Practice effect and test-retest reliability of attentional and executive tests in middle-aged to elderly subjects. *The Clinical Neurologist, 18*, 284-302.

Magnusson, D. (1967). *Test theory*. Reading, MA: Addison-Wesley.

Matías, L. & Zaidspiner, V. (1986). Comparación de los cocientes intelectuales obtenidos en el WISC y en el WISC-R por dos grupos de niños puertorriqueños. Tesis de Maestría inédita, Universidad de Puerto Rico, Recinto de Río Piedras.

Matarazzo, R., Wiens, A., Matarazzo, J., & Manaugh, T. (1997). Test-retest reliability of the WAIS in a normal population. *Journal of Clinical Psychology, 29*, 194-197.

Mc Grew, K. G. & Fanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf ~Gc Cross Battery Assessment.* Massachusetts: Allyn & Bacon.

McPherson, S., Buckwalter, J. G., Tingus, K., Betz, B., & Back, C.

(2000). The Satz- Mogel Short Form of the Weschler Adult Intelligence Scale-Revised: Effects of global mental status and age on test-retest reliability. *Journal of Clinical and Experimental Neuropsychology, 22*, 545-55.

Neyens, L. G. & Aldenkamp, A. P. (1996). Stability of cognitive measures in children of average ability. *Child Neuropsychology, 2*, 161-170.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed). México: McGraw-Hill/Interamericana de México.

Nunnally, J. C. & Bernstein, I. H. (1995). *Teoría psicométrica* (3ra ed.) New York: McGraw-Hill.

Prieto, P. (2004). Introducción a la medición psicológica: psicometría. Accedido el 15 de septiembre de 2005, de http://webpages.ull.es/users/pprieto/material/Psicometria%201.ppt

Raguet, M. L., Campbell, D. A., Berry, D., Schmitt, F. A. & Smith G. T. (1996). Stability of intelligence and intellectual predictors in older persons. *Psychological Assessment, 8*, 154-160.

Sattler, J. M. (2003). *Evaluación infantil, aplicaciones cognitivas.* México: Editorial El Manual Moderno.

Schuerger, J. & Witt, A. (1989). The temporal stability of individually tested intelligence. *Journal of Clinical Psychology*, *45*, 294-302.

Siders, A., Kaufman, A., & Reynolds, C. (2006). Do practice effects on Wechsler's performance subtests relate to children's general ability, memory, learning ability or attention? *Applied Neuropsychology, 4*, 242-250.

Silverstein, A. B. (1991). Reliability of score differences on Weschler's Intelligence Scales. *Journal of Clinical Psychology, 47*, 264-266.

Slate, J. R. & Jones, C. H. (1989). Examiner errors on the WAIS – R: A source of concern. *The Journal of Psychology, 24*, 343-345.

Suzuki, L. A., Meller, P. J., & Ponterotto, J. G. (1996). *Handbook of multicultural assessment: Clinical psychological and educational applications*. San Francisco: Jossey-Bass.

Thompson, A. P. & Molly, D. (1993). The stability of WAIS-I IQ for

16-year-old students retested after 3 and 8 months. *Journal of Clinical Psycholgy*, *49*, 891-898.

Thompson, B. (2006). *Foundations of Behavioral Statistics: An insight-based approach.* New York: Guilford Press.

Tulsky, D. & Zhu, J. (2001). *Escala Weschler de Inteligencia para Adultos III: Manual Técnico.* México: Manual Moderno.

Valcárcel, C. (2000). *Validez y confiabilidad del Inventario para la Clasificación Múltiple de la Inteligencia con una muestra de estudiantes que poseen diferentes rasgos a través de la teoría de inteligencias múltiples*. Disertación doctoral no publicada, Universidad de Puerto Rico, Recinto de Río Piedras, Puerto Rico.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York, TX: The Psychological Corporation.

Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary Scale of Intelligence*. New York: Psychological Corporation.

Wechsler, D. (1968). *Manual de la Escala de Inteligencia Wechsler para Adultos.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1981). *Manual for The Wechsler Adult Intelligence Scale-Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1974). *Manual for The Wechsler Intelligence Scale for Children-Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence - Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Manual for The Wechsler Intelligence Scale for Children, Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1992). *Manual de la Escala de Inteligencia Wechsler para Niños- Revisada de Puerto Rico.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Manual for The Wechsler Adult Intelligence Scale, Third Edition Technical Manual.* San Antonio, TX: The Psychological Corporation.